Journal of Biogeography (J. Biogeogr.) (2009) **36**, 1623–1627

GUEST EDITORIAL

Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling

J. D. Lozier[1]*, P. Aniello[2] and M. J. Hickerson[3]

[1]*University of Illinois, Department of Entomology, Urbana, IL, USA,* [2]*ESRI Database Services, Natural Resource Management, Redlands, CA, USA and* [3]*Queens College, City University of New York, Biology Department, Flushing, NY, USA*

*Correspondence: J.D. Lozier, University of Illinois, Department of Entomology, 320 Morrill Hall, 505 Goodwin Avenue, Urbana, IL 61801, USA. E-mail: jdlozier@life.uiuc.edu

ABSTRACT

The availability of user-friendly software and publicly available biodiversity databases has led to a rapid increase in the use of ecological niche modelling to predict species distributions. A potential source of error in publicly available data that may affect the accuracy of ecological niche models (ENMs), and one that is difficult to correct for, is incorrect (or incomplete) taxonomy. Here we remind researchers of the need for careful evaluation of database records prior to use in modelling, especially when the presence of cryptic species is suspected or many records are based on indirect evidence. To draw attention to this potential problem, we construct ENMs for the North American Sasquatch (i.e. Bigfoot). Specifically, we use a large database of georeferenced putative sightings and footprints for Sasquatch in western North America, demonstrating how convincing environmentally predicted distributions of a taxon's potential range can be generated from questionable site-occurrence data. We compare the distribution of Bigfoot with an ENM for the black bear, *Ursus americanus*, and suggest that many sightings of this cryptozoid may be cases of mistaken identity.

Keywords

Bigfoot, biodiversity informatics, climate change, ecological niche models, North America, Sasquatch, species distributions, *Ursus americanus*.

Ecological niche models (ENMs) and species distribution models have become increasingly popular tools for predicting the geographic ranges of species and have been important for conservation (Kremen *et al.*, 2007), for predicting changes in distribution from past or future climatic events (Hijmans & Graham, 2006), and for investigating patterns of speciation and niche divergence (Wiens & Graham, 2005; Carstens & Richards, 2007; Warren *et al.*, 2008). The basic premise of the ENM approach is to predict the occurrence of species on a landscape from georeferenced site locality data and sets of spatially explicit environmental data layers that are assumed to correlate with the species' range. In many cases, models are based on researchers' own collection data and on detailed knowledge of the taxa being studied, making predictions reasonable depictions of species occurrences given the current modelling technology. However, the increasing availability of locality data in online literature, museum databases and online data portals [e.g. GBIF (http://data.gbif.org/)] is providing unprecedented access to biodiversity data and allowing researchers to greatly expand the deployment of species distribution models and/or ENMs. While the value of publicly available sample locality data is not questioned, the consequent introduction of errors in the accuracy of specimen identity and georeferencing could be problematic for developing ENMs from public data sources (Graham *et al.*, 2004; Soberón & Peterson, 2004). Although georeferencing inaccuracies can be identified in databases from qualitative or quantitative accuracy thresholds (e.g. http://manisnet.org/GeorefGuide.html), poor taxonomy and/or misidentification may be less detectable. This issue may be particularly problematic, for example, with cryptic species or subspecies that are morphologically similar but may have very distinct ecological requirements and geographic distributions, or for those data sources that contain indirect observations rather than references only to physical specimens.

In our own attempts at implementing ENMs we have encountered probable misidentifications in biodiversity records for a number of species, and expect that many researchers have had similar issues. To demonstrate the potential for generating dubious yet visually convincing distributions from publicly available data we use ENMs to predict the range not of misidentified cryptic species, but of a crypto-zoological species – the North American Sasquatch, or Bigfoot. Supposedly,

© 2009 Blackwell Publishing Ltd

doi:10.1111/j.1365-2699.2009.02152.x

Sasquatch belongs to a large primate lineage descended from the extinct Asian species *Gigantopithicus blacki*, but see Milinkovich *et al.* (2004) and Coltman & Davis (2005) for phylogenetic analyses indicating possible membership in the ungulate clade. Sasquatch is regularly reported in forested lands of western North America, as well as being considered a significant indigenous American and western North American folk legend (Meldrum, 2007), however the existence of this creature has never been verified with a typed specimen.

We present ENMs for Bigfoot in western North America based on a repository of (1) putative sightings and auditory detections ($n = 551$), and (2) footprint measurements ($n = 95$) collected from 1944 to 2005 (Fig. 1). This data set was taken from a collection of reported Bigfoot encounters archived by the Bigfoot Field Researchers Organization (BFRO; http://www.bfro.net/news/google_earth.asp). The reports generally consisted of a description of the event and where it occurred. The reports used were filtered to eliminate spurious points by carefully examining event descriptions prior to incorporation in the present study. The events were assigned geographic coordinates by matching descriptions of event locations to actual locations on USGS quad maps and commercially available atlases.

Figure 1 Map of Bigfoot encounters from Washington, Oregon and California used in the analyses. Points represent visual/auditory detection, and foot symbols represent coordinates where footprint data were available. Shading indicates topography, with lighter values representing lower elevations.

ENMs were constructed using the maximum entropy niche modelling approach implemented in the software MAXENT v3.1 (Phillips *et al.*, 2006), with 80% of the data used in training and 20% retained as test points. Environmental data layers were constructed for the 19 BIOCLIM variables (at 5-arcminute resolution) in the WORLDCLIM data set (Hijmans *et al.*, 2005). To reduce the number of bioclimatic variables in order to minimize model overfitting, we extracted climate data for 5000 random points sampled from the geographic extent of our study and calculated correlations between each variable for these points. For pairs with a correlation coefficient > |0.80|, one variable was selected. This resulted in a set of nine variables: (1) annual mean temperature, (2) mean diurnal range, (3) isothermality, (4) temperature annual range, (5) mean temperature of wettest quarter, (6) mean temperature of driest quarter, (7) precipitation seasonality, (8) precipitation of warmest quarter, and (9) precipitation of coldest quarter.

MAXENT appeared to perform well in our analysis, according to the area under the receiver operating characteristic curve (AUC) and threshold-based evaluation methods (Phillips *et al.*, 2006). The ENM had an AUC of 0.983 and strongly rejected the hypothesis that test points are predicted no better than by a random prediction for all thresholds implemented in MAXENT. No locality points fell outside the predicted distribution, with the exception of a single observation from Imperial, California. In general, the ENM shows that Bigfoot should be broadly distributed in western North America, with a range comprising western North American mountain ranges such as the Sierra Nevada Mountains, the Cascades, the Blue Mountains, the southern Selkirk Mountains, and the Coastal Range of the Pacific Northwest. Based on jackknife analyses for models including each variable alone, 'precipitation of the coldest quarter' was the bioclimatic variable that contributed most to the ENM, followed by 'temperature annual range', 'mean temperature of the wettest quarter', and 'mean temperature of the driest quarter'. An ENM produced using footprint data alone was highly similar to that for all sighting data (not shown).

It is expected that species distributions are likely to be altered under global warming, and a number of studies have used environmental layers derived from climate-change models to project contemporary distributions into the future using ENMs (Hijmans & Graham, 2006; Pearson *et al.*, 2006; Loarie *et al.*, 2008). Despite potential weaknesses in such approaches (Pearson *et al.*, 2006), we were interested in examining the potential ramifications of climate change on hypothetical remnant Sasquatch populations and to predict how the frequency of sightings might change in the future. We projected ENMs generated from the WORLDCLIM data onto bioclimatic layers simulated for a doubling of atmospheric CO_2 (http://www.diva-gis.org/climate.htm). As expected for montane organisms, the model predicts Bigfoot to abandon

© 2009 Blackwell Publishing Ltd

13652699, 2009, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2009.02152.x, Wiley Online Library on [14/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

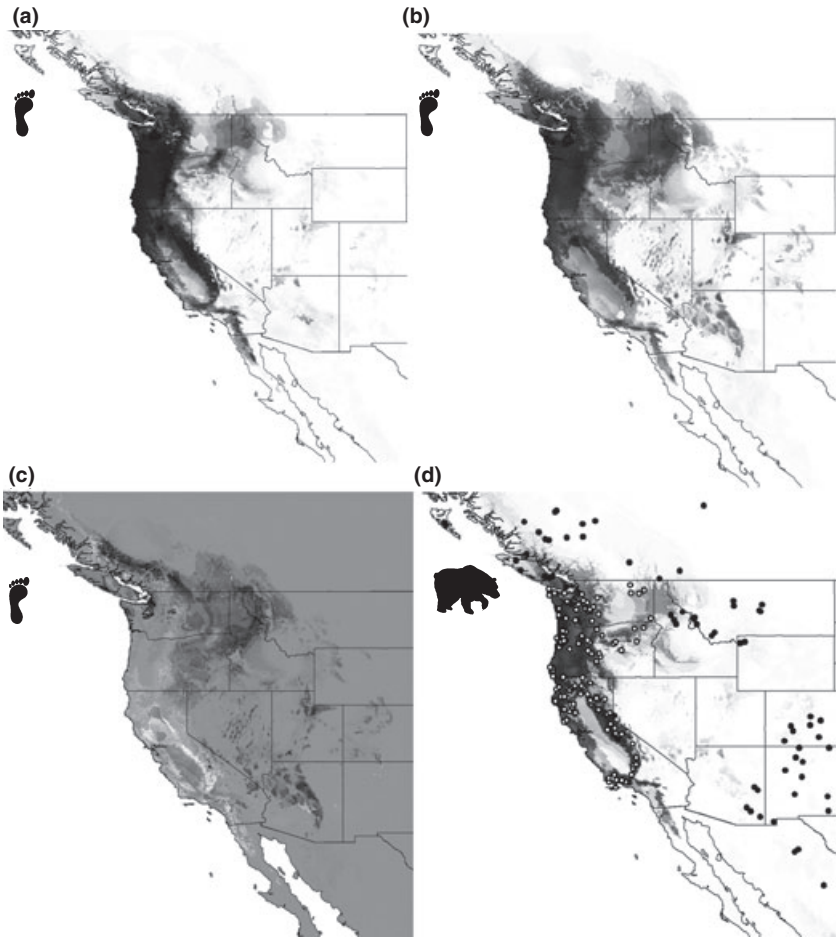

Figure 2 Predicted distributions of Bigfoot constructed from all available encounter data using MAXENT (a) for the present climate and (b) under a possible climate-change scenario involving a doubling of atmospheric CO_2 levels. Results are presented for logistic probabilities of occurrence ranging continuously from low (white) to high (black). Differences between (a) and (b) are shown in (c), with whiter values reflecting a decline in logistic probability of occurrence under climate change, darker values reflecting a gain, and grey reflecting no change. A predicted distribution of *Ursus americanus* in western North America under a present-day climate is also shown (d). White points indicate sampling localities in California, Oregon and Washington taken from GBIF (n = 113 for training, 28 for testing; compare with Fig. 1) used for the MAXENT model with shading as in (a) and (b); black points indicate additional known records not included in the model.

lower altitudes and also to lose habitat in coastal regions (Fig. 2b, c). However, this loss of habitat should be compensated by a large potential gain in the northern part of the Sasquatch range and in several other montane areas (e.g. Arizona, Nevada, Utah), should such areas remain undisturbed by human activity in the near future (Fig. 2c). Thus, given our model and available data, we might expect Bigfoot sightings to increase in frequency in northern latitudes and at higher elevations over the coming years.

Notably, the predicted distribution of Sasquatch (Fig. 2a) appears similar to that which might be expected for other large mammals of western North America, including the American black bear (*Ursus americanus* Pallas, 1780), sightings of which are thought to be sometimes confused as Bigfoot encounters (Meldrum, 2007). We sampled records of *U. americanus* from the same states as considered for Bigfoot (California, Oregon and Washington) from GBIF. We selected records for which physical specimens were available, allowing us to reasonably assume that black bears were not misidentified, and for which site localities could be georeferenced to a named place at minimum. Although this level of geographical accuracy may not be ideal for many ENM applications, it should be sufficient for our purposes here. We performed MAXENT runs as described above (present climate only). The model (Fig. 2d) performed well according to AUC (both test and training AUC > 0.98) and all threshold statistics (all $P < 10^{-47}$). However, visualizing all black bear records from GBIF reveals that limiting model training to locations within California, Oregon and Washington leads to an under-prediction of the known distribution of *U. americanus*, particularly in New Mexico, Colorado, and parts of Canada and Mexico (Fig. 2d). Although this suggests some methodological limitations for the distribution models as implemented here, these other locality points may belong to different subspecies that experience unique environmental conditions in inland regions relative to the more coastally distributed specimens used in our analysis (Larivière, 2001). Our ENM prediction does in fact appear quite similar to the distributions of the subspecies *U. a. altifrontalis* and *U. a. californiensis* (Larivière, 2001), supporting this possibility. In any event, for our comparisons here, any under-prediction should have little consequence since we expect such limitations to be shared by the Sasquatch model, given the similar distribution of locality points.

The general similarities between distributions of the two 'species' are clear (Fig. 2a, d), despite the much smaller number of available black bear coordinates. Furthermore, the exact same bioclimatic variables (see above) contributed most to the ENM when evaluated using MAXENT's variable jackknifing procedure. We used the *I*-statistic (Warren *et al.*, 2008) to quantify the degree of similarity between the two ENMs

© 2009 Blackwell Publishing Ltd

13652699, 2009, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2009.02152.x, Wiley Online Library on [14/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

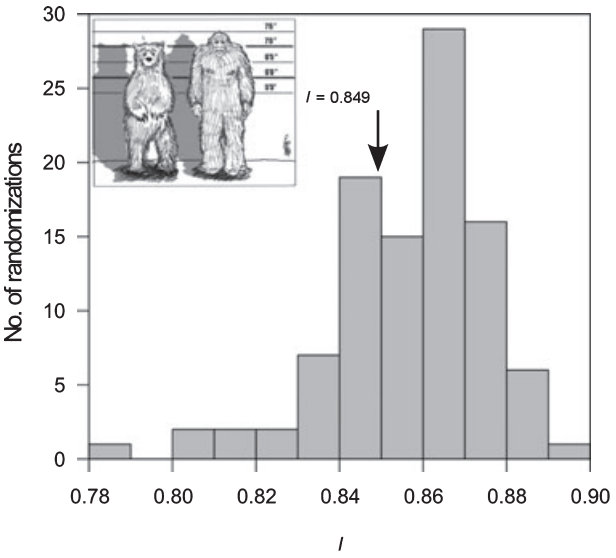

Figure 3 A null distribution of the ecological niche model (ENM) overlap statistic *I* created from MAXENT runs for 100 randomizations of localities between Bigfoot and black bear (*Ursus americanus*) data sets. The arrow shows the *I* value observed (0.849) for the actual data sets for MAXENT runs using all locality points (i.e. no test points).

using the program ENMTOOLS. The observed value of $I = 0.849$ indeed indicates a high degree of overlap, and falls well within the null distribution generated from MAXENT runs for 100 randomizations of Bigfoot and black bear coordinates (Fig. 3; $P <$ observed $= 0.32$). Thus, the two 'species' do not demonstrate significant niche differentiation with respect to the selected bioclimatic variables. Although it is possible that Sasquatch and *U. americanus* share such remarkably similar bioclimatic requirements, we nonetheless suspect that many Bigfoot sightings are, in fact, of black bears.

CONCLUSIONS

We were stimulated to write this piece as a tongue-in-cheek response to the increasing prevalence of ENMs in the literature and in papers presented at professional meetings. As in any rapidly developing field with the promise of exciting applications, there is the potential for the empirical acceptance of new approaches to outpace conceptual understanding. The point of this paper has been to point out how very sensible-looking, well-performing (based on AUC and threshold tests) ENMs can be constructed from questionable observation data. The fact that a model can be fit to any random assemblage of locality points is, of course, not surprising, but in cases where the distribution of these points appears to be organized, records may be less carefully scrutinized for errors. For instance, the Sasquatch sightings shown in Fig. 1 occur in areas where we might expect to observe such an organism based on preconceived notions, but we may be seriously biasing our inference of the distribution if many (or all) of the records represent misidentified black bear sightings. Although beyond the scope of this piece, it would be interesting for future theoretical studies to explore the effects of including misidentified specimens on ENMs, either using simulations or empirical data from closely related species.

We stress that our aim here is not to disparage the value of literature records or public specimen databases, or to discourage the use of species distribution modelling, but rather to encourage careful scrutiny of specimen records prior to their use in ENMs. Wherever feasible we recommend that researchers should attempt to access specimens with suspect records in order to check their validity prior to application in ENMs. In addition, determinations made by taxonomic experts might be favoured over those by unrecognized individuals, all other information being equal. The addition of uncertainty codes like those used for georeferencing would greatly assist users in filtering databases by taxonomic accuracy. Finally, as museums begin to develop databases of collections, all effort should be made to ensure taxonomic accuracy prior to releasing information to the public and to update these databases as new research accumulates for difficult species groups (Graham *et al.*, 2004; Soberón & Peterson, 2004; Chapman, 2005). Together, all of these recommendations suggest that as specimen data increasingly become easily available, the need for well-trained taxonomists and systematists has never been greater.

Finally, because our aim was to keep this commentary light-hearted, we have not addressed many other potential sources of error in ENMs that have been discussed in the literature. A number of recent papers have dealt with the issues of environmental variable selection, modelling methods, model evaluation, projecting models into the past and future, biotic interactions, and the important consideration of potential versus realized distributions/niches, and we refer readers to these excellent resources (e.g. Soberón & Peterson, 2004, 2005; Araújo & Guisan, 2006; Pearson *et al.*, 2006; Peterson, 2006; Araújo & Luoto, 2007; Guisan *et al.*, 2007; Jiménez-Valverde *et al.*, 2008; Phillips & Dudík, 2008). Although our possible underestimate of the known black bear distribution demonstrates that such methodological limitations may sometimes be problematic when occurrence data are limited to a subset of a species' range, we believe that ENMs will remain important tools for understanding the spatial distributions of biodiversity. We simply urge researchers to take care in evaluating the suitability of data sets prior to their use in such distribution modelling approaches.

ACKNOWLEDGEMENTS

We greatly thank C. Zartman and E. Gomez for providing illustrations, W. Monahan for initial discussions, and two anonymous reviewers for making useful suggestions. We also very much thank R. Whittaker and R. Pearson for handling the manuscript, Milinkovich *et al.* (2004) and Coltman & Davis (2005) for their pioneering work and inspiration for this paper, and whoever invented April Fools' Day.

© 2009 Blackwell Publishing Ltd

13652699, 2009, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2009.02152.x, Wiley Online Library on [14/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **10,** 1677–1688.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16,** 743–753.

Carstens, B.C. & Richards, C.L. (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61,** 1439–1454.

Chapman, A.D. (2005) *Principles of data quality, version 1.0*. Report for the Global Biodiversity Information Facility, Copenhagen.

Coltman, D. & Davis, C. (2005) Molecular cryptozoology meets the Sasquatch. *Trends in Ecology and Evolution*, **21,** 60–61.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19,** 497–503.

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. & the NCEAS Species Distribution Modelling Group (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13,** 332–340.

Hijmans, R.J. & Graham, C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12,** 1–10.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25,** 1965–1978.

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14,** 885–890.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S.J., Thomas, C.D., Beentje, H., Dransfield, J., Fisher, B.L., Glaw, F., Good, T.C., Harper, G.J., Hijmans, R.J., Lees, D.C., Louis ., E., Jr, Nussbaum, R.A., Raxworthy, C.J., Razafimpahanana, A., Schatz, G.E., Vences, M., Vieites, D.R., Wright, P.C. & Zjhra, M.L. (2007) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320,** 222–226.

Larivière, S. (2001) *Ursus americanus*. *Mammalian Species*, **647,** 1–11.

Loarie, S.R., Carter, B.E., Hayhoe, K., McMahon, S., Moe, R., Knight, C.A. & Ackerly, D.D. (2008) Climate change and the future of California's endemic flora. *PLoS ONE*, **3,** e2502.

Meldrum, J. (2007) *Sasquatch: legend meets science*. Forge Books, New York, NY.

Milinkovich, M.C., Caccone, A. & Amato, G. (2004) Molecular phylogenetic analyses indicate extensive morphological convergence between the ''yeti'' and primates. *Molecular Phylogenetics and Evolution*, **31,** 1–3.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33,** 1704–1711.

Peterson, A.T. (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, **3,** 59–72.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31,** 161–175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190,** 231–259.

Soberón, J. & Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359,** 689–698.

Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2,** 1–10.

Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62,** 2868–2883.

Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Reviews of Ecology, Evolution, and Systematics*, **36,** 519–539.

Editor: Robert Whittaker

© 2009 Blackwell Publishing Ltd

13652699, 2009, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2009.02152.x, Wiley Online Library on [14/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License